Measuring the quality of judgement and decision-making in nursing

Dawn Dowding BSc PhD RGN
Senior Lecturer in Clinical Decision Making, Department of Health Sciences and the Hull York Medical School, University of York, York, UK

and Carl Thompson BA RN DPhil
Research Fellow, Department of Health Sciences, University of York, York, UK

Submitted for publication 16 January 2003
Accepted for publication 18 June 2003

Correspondence:
Dawn Dowding,
Department of Health Sciences and HYMS,
University of York,
Seebohm Rowntree Building,
Heslington,
York YO10 5DD,
UK.
E-mail: dawn.dowding@hyms.ac.uk

DOWDING D. & THOMPSON C. (2003) *Journal of Advanced Nursing* **44**(1), 49–57

Measuring the quality of judgement and decision-making in nursing

Aim. This paper discusses measurement of the quality of judgement and decision-making in nursing research. It examines theoretical and research issues surrounding how to measure judgement accuracy as a component of evaluating decision-making in nursing practice.

Discussion. Judgement accuracy is discussed with reference to different methods of measurement, including comparing judgements with independent criteria and inter-judge approaches. Existing research on how judgement accuracy has been measured in nursing practice is examined. Evaluation of decisions is then discussed, including consideration of the process of decision-making and evaluating decision outcomes. Finally, existing research on decision-making in nursing is assessed and the strengths and limitations of different types of measurement discussed.

Conclusion. We suggests that researchers examining the quality of judgement and decision-making in nursing need to be aware of both the strengths and limitations of existing methods of measurement. We also suggest that researchers need to use a number of different methods, including normative approaches such as Bayes' Theorem and Subjective Expected Utility Theory.

Keywords: clinical decision-making, clinical judgement, nursing

Introduction

Nurses and midwives, like all health care professionals, have had to adapt to the increased emphasis on evidence-based health care decisions in clinical practice. Recent policy changes and trends in professional development mean that nurses need to recognize that the decisions they make have a significant impact on health care outcomes and patients' experiences (Department of Health 2000). Whilst knowledge of the means of generating the evidence for practice is relatively well developed, our understanding of how nurses use this knowledge in their clinical decisions and judgements is comparatively sparse. The increasingly blurred boundaries between health care professions mean that nurses' judgements and decisions have more potential than ever to impact on patients' lives and experiences. How we measure the accuracy or 'goodness' of nurses' judgements and decisions is therefore of prime importance.

The purpose of this paper is to discuss ways of measuring the quality of judgement and decision-making in nursing. It first discusses theoretical approaches to measuring judgement accuracy and evaluating decision-making, before analysing how this has been approached in the nursing literature. Although judgement and decision-making are closely linked, we have dealt with the two concepts separately. This is because they generate separate cognitive demands, and pose unique and distinct challenges for researchers seeking to describe and evaluate them.

Literature search

As a starting point for the paper a literature search was carried out in CINAHL (1982–November 2002) and Medline (1966–2002), using key word searches and limited by language and type of publication (Table 1). Studies were selected on the basis that they were reports of research that used some form of measurement to examine the accuracy of judgement or evaluate decisions in nursing, with an explicit comparison of judgements or decisions with some form of criterion. The 17 studies that were identified form the basis of the discussion of measurement issues.

Judgements and decisions

One way of looking at judgement is as 'an assessment between alternatives' (Dowie 1993). Maule (2001) suggests that the process of judgement involves the integration of different aspects of information about a person, object or situation to arrive at an overall evaluation. Therefore the central question for those researching judgement is, 'How do nurses use different types of clinical information about the patient (how they look, their vital sign readings, their medical condition, their behaviour) to arrive at a judgement of the patient's current health status?'. This field of research also considers how individuals make predictions about risk, such as how likely it is that a patient will develop a pressure ulcer.

As opposed to the assessment of information, decisions have been defined as 'a choice between alternatives' (Dowie 1993). The research emphasis here is *how* people choose particular courses of action, especially in situations of uncertainty where the consequences of their actions are unknown (Goldstein & Hogarth 1997). In nursing an example of a clinical decision might be the choice of wound care product for use with an individual patient.

It is important to distinguish between these two concepts. For instance a nurse or midwife may make an accurate judgement (e.g. 'this patient is in extreme pain') but then choose a poor quality action, i.e. make a 'bad' decision (e.g. I'll give them a heat pad). Equally they may make a poor judgement ('this person's chest pain is due to indigestion', when actually it is cardiac pain) but make a good decision on the basis of the poor judgement ('I'll give them medication for indigestion'). In this instance the nurse made a perfectly reasonable decision given the information available: the wrong diagnosis. Of course, what most clinicians strive for is the synergistic state of both judgements and decisions being of 'good' quality. For example, the individual's current situation is assessed accurately and appropriate action taken on the basis of that assessment ('this person has extreme chest pain, so I will give them morphine and do an ECG').

Table 1 Search strategy

1. Judgement OR Judgment
2. Decision
3. 1 OR 2
4. Nurse OR Nursing
5. 3 AND 4
6. Limited to RESEARCH
7. Limited to English language

Quality of judgement

Accuracy as quality of judgement

Hastie and Rasinski (1988) define judgement accuracy as having no errors, being correct, or deviating only slightly but within acceptable limits from a standard. They go on to suggest that analysing the accuracy of judgements requires three elements:

- the judgement, response or assertion under consideration,
- the standard or criterion of truth and
- the rule specifying the correspondence relation between the judgement and the criterion.

Cooksey (1996) argues that an individual's ability to make correct judgements is a function of three things:

- how predictable the world is,
- how well the judges know the world (i.e. their knowledge base) and
- how consistently judges apply their knowledge.

One theoretical framework uniting these assumptions and providing a way of measuring judgement accuracy is social judgement theory (SJT). Central to this is the idea that an individual's judgement relate to the reality of a social environment and can be conceptualized as a 'lens'. Moreover, this idea of a lens (Figure 1) can be used to model the ways in which various forms of information relate to the 'reality' of a given judgement environment and how individuals use information to arrive at their judgements. The model suggests that the 'ecological' situation (what is wrong with the patient for instance) is on the left hand side of the model. There are a variety of different pieces of information (cues) that are probabilistically related to this ecological situation (e.g. the patient's signs and symptoms), with different importance or weight attached to them. The judge uses these cues to make a judgement (the right hand side of the model). The judgement outcome is a function of how the cues have been used. If the cues are weighted in the same way by the judge as they are linked to the ecological situation, then the judgement will be more accurate. If the judge weights the cues differently, then

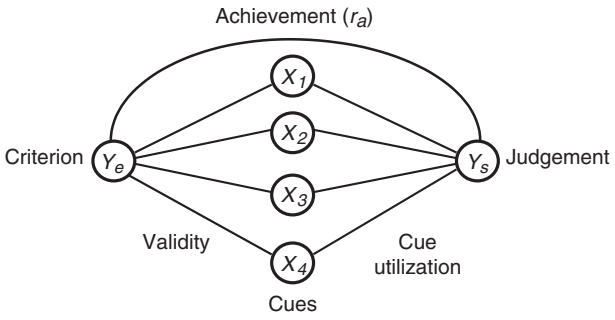

Figure 1 Brunswik's lens model (Hammond 1975, p. 274). Y_e is the criterion or actual patient state. X_1-X_n are the cues which are related to that patient state. They may vary in importance or 'weight' with regard to their relationship to the patient state. Y_s is the persons judgement about the criterion or patient state. They may use the cues X_n in a different way to how they are actually related to Y_e. r_a is the persons achievement.

their judgement will not reflect the ecological situation and be less accurate. Regression techniques are used to derive a statistical equation or algorithm revealing how much weight is attached to each item of information related to the ecological situation, or used in the judgement (Cooksey 1996).

Measuring judgement quality

Criterion and basic logic approaches

Basic logic (Hastie & Rasinski 1988) approaches to measuring judgement rely on the definition of the judgement task, measuring the criterion, independently measuring the subject's judgement and comparing the judgement with the criterion. This can be done in two ways: through social judgement approaches or probabilistic methods.

Social judgement approaches use the relationship between the information and outcomes of interest as the basis for establishing the criterion (the left hand side of the lens). The way this is typically represented is through the use of linear models, developed through regressing information cues onto the criterion (Engel *et al.* 1990). Linear models constructed in this way give a measure of task predictability, providing an upper limit on how well we can then expect an individual judge to perform when they are making judgements within this particular context (Engel *et al.* 1990, Cooksey 1996, Dawes 2000, Harries & Harries 2001). For example, Moore *et al.* (1996) suggest that only 80% of all falls in older patients are predictable. In this instance a criterion of being able to predict 80% of the falls that occur would indicate accurate judgement. Individuals are then assessed against a judgement task based on the modelled environmental reality. The assessment provides a picture of the information they use and the weightings they attach to the information used in reaching their judgement (again using linear regression models). Despite arguments that actual cognitive processes may be more complex than simple linear equations, linear models have consistently been shown to provide an accurate description of many judgement processes, including that most complex of health care activities, diagnosis (de Dombal 1988, Hastie & Dawes 2001).

Probabilistic methods can also be used to assess accuracy of judgement. For instance, the likelihood of an individual having a particular condition can be calculated from existing data (Jungermann 2000) and compared with a judge's evaluation of the likelihood of an individual having the condition. Normative approaches such as Bayes' theorem (Hastie & Dawes 2001) can provide the probability of an individual having a particular medical diagnosis, given the prevalence of the disease in the population and the information provided by test results. The criterion in this instance would be the predictions made by the normative model (Bayes' theorem) and performance would be compared against this. Probabilistic approaches rely on the availability of data sets that can provide the necessary frequency data with which to evaluate the accuracy of judgements made by individuals.

Inter-judge comparisons

An alternative approach to measuring the accuracy of judgement is by comparing the judgements of separate judges and examining them for agreement. The assumption here is that, if they disagree, one of them must be in error (Hastie & Rasinski 1988). There is no independent criterion in this approach and so there is always the possibility that *both* individuals may be in error. Inter-judge approaches can lead to systematic measurement errors, and so it is clearly advantageous for those considering researching the accuracy of clinical judgement to use independent criterion approaches. We now turn to how these approaches have been employed in looking at the measurement of judgement by nurses.

Measuring judgements in nursing practice

Basic logic approaches

Studies within this approach have used two types of criterion measure against which to examine nurses' judgements, one being actual patient outcome/condition and the other being a measurement of patient status. Seven studies have been identified that use the actual patient condition or outcome as a measure against which nurses' judgements have been

assessed (Kruse *et al.* 1988, Rosenthal *et al.* 1992, Moore *et al.* 1996, VandenBosch *et al.* 1996, Marsden 2000, Moyer *et al.* 2000, Allen-Davis *et al.* 2002) and three other measurements of patient status (McDonald *et al.* 1999, Reid & Chappell 2000, O'Brien *et al.* 2001).

Four of the studies examine nurses' diagnostic accuracy, comparing their performance to actual patient diagnoses established via laboratory tests (Rosenthal *et al.* 1992, Moyer *et al.* 2000, Allen-Davis *et al.* 2002) or diagnosis determined from patient notes (Marsden 2000). Allen-Davis *et al.* (2002) examined nurses' ability to diagnose vulvovaginal complaints over the telephone compared with actual diagnosis determined by laboratory tests. Rosenthal *et al.* (1992) compared nurse practitioners' probability estimates of whether women had chlamydial infection following an examination to results from a swab analysed by a virology laboratory. Moyer *et al.* (2000) asked paediatric nurse practitioners to provide judgements on the presence of neonatal jaundice on the basis of a physical examination, and compared them with serum bilirubin concentration measurements. All of these studies were carried out in the United States of America, and it is not clear whether nurses would normally be expected to make such diagnoses or assessments without the additional information provided by laboratory tests. Rosenthal *et al.* (1992) used linear modelling to identify the weighting of clinical cues that were associated with actual infection, and the weighting of the cues that nurses used to make their probability judgements. This allowed them to identify potential reasons for judgement error in the nurses they studied, including inconsistent use of information and over/underestimation of the importance of some clinical cues.

Marsden (2000) compared the provisional telephone diagnosis of nurse practitioners practising in an eye department to patients' subsequent diagnosis. This study was carried out in the United Kingdom, using a retrospective secondary data analysis of patient triage records, hospital records and subsequent visits, with the researcher interpreting whether or not the diagnosis made by the nurse was appropriate or not.

The strength of these studies is that they all use patient outcomes as a measure against which to compare nurses' judgements. They also have outcome data for a reasonably large sample of patients [492 women (Rosenthal *et al.* 1992), 253 women (Allen-Davis *et al.* 2002), 122 healthy infants (Moyer *et al.* 2000), 461 patients (Marsden 2000)]. However, the number of practitioners involved in the studies is comparatively small (four in Rosenthal *et al.* 1992, seven in Marsden 2000) or not stated at all (Moyer *et al.* 2000, Allen-Davis *et al.* 2002). For three of the studies it is unclear whether these tasks are ones which nurses would normally be expected to carry out (Rosenthal *et al.* 1992, Moyer *et al.* 2000, Allen-Davis *et al.* 2002) and for the other there are questions about relying on patient notes as the source of data and researcher interpretation of the notes as a way of measuring judgement accuracy (Marsden 2000).

The remaining three studies all compare nurses' risk predictions of patient status with whether or not the patient developed the outcome of interest. Moore *et al.* (1996) asked nurses to estimate, at the time of admission to hospital, a patient's risk of falling (low, medium or high) and then compared this to whether patients actually fell or not, identified through hospital incident reports and audit forms. Kruse *et al.* (1988) asked nurses to estimate, at the time of admission to an intensive care unit, the risk of a patient dying whilst in hospital, and then looked at patient outcomes (whether they survived or died). They placed probability judgements into two categories to carry out the comparison, with judgements above 50% equating to 'die' and below to 'survive'. VandenBosch *et al.* (1996) asked nurses to provide a yes/no response about whether or not they thought that a patient would develop a pressure ulcer whilst in hospital, then followed patients up to see if they subsequently developed ulcers.

All of these studies were carried out in the USA, with the number of assessments and patients taking part in the studies varying from 39 patients (187 pairs of judgements) (Moore *et al.* 1996), to 103 patients (VandenBosch *et al.* 1996) and 366 patients (Kruse *et al.* 1988). None of the studies provide details of the number of nurses who provided risk judgements about the patients. Again, one of the strengths of these studies is that they are using patient outcomes as a measure of judgement accuracy. However, when examining risk prediction accuracy there are two issues that also need to be taken into account, and that were not addressed by any of these studies. The first is the predictability of the condition in the first place. Moore *et al.* (1996) suggest that only 80% of falls can be predicted (but fail to provide evidence for this figure), and neither Kruse *et al.* (1988) nor VandenBosch *et al.* (1996) give any insight into what a reasonable level of prediction might be in these circumstances. The second issue concerns the effect of preventative action when an individual is identified as being at risk (Papanikolaou *et al.* 2002). When using patient outcomes from a situation where the task is prediction of risk, it is unclear what effect interventions taken to reduce that risk may have had on the outcome. It is therefore difficult to assess how accurate initial risk predictions made by practitioners may have been.

Of the three studies that use other measurements of patient status, two use comparisons with a patient-completed measurement tool (McDonald *et al.* 1999, O'Brien *et al.* 2001) and

one a researcher-completed measurement tool (Reid & Chappell 2000). McDonald *et al.* (1999) asked patients to complete the Zung self-rating depression scale, and compared these levels of depression with nurses' ratings of levels of depression. Similarly O'Brien *et al.* (2001) asked patients to rate their own anxiety using the Spielberger State Anxiety Index, and compared these ratings to nurses' assessments as recorded in medical records. Reid and Chappell (2000) filled out the revised multi-focus assessment scale on the basis of information they obtained from interviews with patients with dementia and their families, and then compared the results to assessments of level of dementia obtained via Directors of Nursing.

The methods used in these studies are different from those measuring patient outcomes directly, in that they are interested in patient characteristics that are difficult to determine without some form of assessment scale. All of the studies assume that the patient measurements are reliable and valid. They also vary considerably in how they measure nurses' judgements using direct assessment (McDonald *et al.* 1999); the assesment of directors of nursing who may or may not have consulted with other nursing staff (Reid & Chappell 2000); or what was written in the patient's medical notes (O'Brien *et al.* 2001). Sample sizes vary from 1109 patients and 40 nurses (McDonald *et al.* 1999), through 510 patients (Reid & Chappell 2000) to 40 patients (O'Brien *et al.* 2001). The main issue with these types of studies is whether the criterion used to measure patient status is an accurate representation of physical or mental state. Two of the studies would have also benefited from measuring nurses' judgements directly, rather than depending on written sources or Directors of Nursing.

Inter-judge comparisons

Studies that have used inter-judge comparisons have primarily used expert panel consensus as the criterion against which nurses' judgements have been compared (Aspinall 1979, Letourneau & Jensen 1998, Gould *et al.* 2001, Reischman & Yarandi 2002), although two use the judgements of other nurses (Westfall *et al.* 1986, Melchior-MacDougall & Lander 1995) and one uses group consensus obtained via a nominal group technique (Zou *et al.* 1998).

All of the studies using expert panel consensus as a gold standard for nurses' judgements also use written case simulations as a method of data collection. Gould *et al.* (2001) used four written simulations developed from case histories of patients to examine nurses' judgements about pressure ulcer risk. They used an expert panel to achieve consensus on the level of risk that each simulation represented, before asking 236 nurses to make a judgement on the level of risk. Reischman and Yarandi (2002) similarly used written patient simulations that had been reviewed by an expert panel to compare how 23 expert and 23 novice nurses used information to reach diagnoses for patients in cardiovascular critical care. However, they fail to provide detail on how the diagnostic content was initially obtained.

Aspinall (1979) gave nurses a written case study and asked them to list the possible causes of the patient's behaviour. An expert panel were used to identify which of the 18 possible disease states (diagnoses) could be a 'correct' diagnosis for the patient. The list of six 'correct' diagnoses was then used to calculate an index of diagnostic accuracy, which was used to assess the performance of the 90 nurses who took part in the study. Letourneau and Jensen (1998) used three case studies previously developed by Melchior-MacDougall and Lander (1995) to examine 163 home care nurses' accuracy in wound staging. They revalidated the case studies using an expert panel to identify what stage of the wound represented in the case studies.

As in the studies that use an expert panel to establish the criterion for judgement, the two studies that use other nurses to establish criterion also use simulations as a data collection method. Melchior-MacDougall and Lander (1995) used three case histories based on actual patient cases, and asked 94 home care nurses to stage the wound. Local practitioners assessed the accuracy of their judgements. Westfall *et al.* (1986) used patient simulations based on actual patient cases to examine 43 nursing students' and nurses' diagnostic reasoning, assessing whether or not they generated accurate hypotheses or diagnoses for the patient cases. The research team and nurses providing the information for the simulations determined the criterion for accuracy *a priori*. In contrast, Zou *et al.* (1998) asked nurses to assess patients for the presence of delirium before they took part in a multi-professional group that established a consensus on 87 patients' actual diagnoses.

Using simulations has strengths as a method of data collection, in that all subjects see the same cases and so judgements can be compared easily. However, there are issues of representativeness, in that it is uncertain how close to reality such simulations are, and whether or not judgements made in these environments are representative of those made in 'real life' (Lamond *et al.* 1996). Expert panels can be a useful mechanism for establishing a criterion for performance, especially in areas such as pressure ulcer risk, where considering actual patient outcomes is not without its limitations. However, one has to be aware that expert consensus by its nature means that some individuals may have compromised their own judgements to reach

agreement – which may mean that it does not represent the 'best' judgement in every situation. This is similarly the case with studies that use a nominal group technique to establish consensus. Establishing accuracy using a form of peer review may provide an insight into what is considered 'good' practice in that particular context or setting, but this limits the standard to existing performance rather than trying to identify whether new levels of achievement are possible (Gonzalez 2001).

The 'goodness' of decisions

What makes a good decision?

Defining decisions as good or bad is problematic, mainly because nurses operate in an environment that is characterized by uncertainty (Buckingham & Adams 2000). Baron (2000) suggests that the best decisions are those that yield the best consequences for achieving people's goals. However, evaluating decisions by their outcomes can be misleading as a measure of quality (Pauker & Pauker 1999, Sox 1999), as the outcome may have occurred by chance, although the decision was the 'best' one for the individual at the time. Another alternative to measuring decisions is to evaluate the process by which the decision has been made (Pauker & Pauker 1999). However, this is also problematic, as it ignores the outcome of the decision and raises the issue of what makes a 'good' decision process. Hastie and Dawes (2001) suggest that good decisions are those in which the process follows the laws of logic and probability theory. Normative theories, such as subjective expected utility theory (SEUT) could be considered an optimum process by which to make decisions. SEUT explicitly takes the decision-maker's values or beliefs into account (Dowie 1993), and follows a rational process that includes the probability of various outcomes occurring, before identifying the optimum decision for that individual (Tavakoli *et al.* 1999).

Evaluating decisions

There are a number of ways in which decisions can be evaluated, including examining the outcome of the decision, the decision process itself, or comparing decisions either to some form of normative model or to another individual. As has already been highlighted, evaluating the outcome of a decision can be problematic as it could be determined by chance events (Pauker & Pauker 1999, Sox 1999). However, Sox (1999) suggests it may be reasonable to evaluate the decision-maker with reference to the outcomes of many decisions (as one can then consider average performance). If the decision process is being evaluated as a measure of quality, then an idea of what should be included in that process is necessary. Pauker and Pauker (1999) suggest that the process should be explicit, define the decision problem, identify the goals of the decision-maker, specify the consequences and relative values of the outcomes for each option, examine the trade-offs between each of the strategies and include all relevant parties in the process.

When comparing decisions to a normative model such as SEUT, it is assumed that the normative model is following a 'good' decision process. If the choice the decision-maker has made is evaluated against the recommendation of the normative model and they agree, then it could be considered 'good'. However, it is necessary to be aware that the agreement may have resulted by chance (Sox 1999). If decisions are compared between decision-makers, perhaps with some form of expert performance as a gold standard, then what is being suggested is that the expert is using what is considered to be a 'good' process (Lipshitz *et al.* 2001). However, similar concerns to those raised in the section on judgement quality are attached to using inter-judge comparisons as a means of assessing decision quality.

Evaluating decisions in nursing practice

Studies that have attempted to evaluate the decision-making of nurses have typically used one main strategy; that of comparing nurses' decisions to a 'gold standard' either determined through expert consensus, peer evaluation or what happened to the patient.

Three studies have examined the appropriateness of nurse triage decision-making (Leprohon & Patel 1995, Considine *et al.* 2000, Quinn *et al.* 2000). Considine *et al.* (2000) used 10 scenarios based on patient cases to investigate the appropriateness of triage decisions in 31 nurses. Triage decisions were compared with those established via expert panel consensus. In a different study, Leprohon and Patel (1995) examined the case records of telephone triage phone calls and followed up patients via their notes. This information was then given to experts, who were asked to judge what the optimal decision was for each call, and nurses' decisions were compared with this. This study examined 34 nurses receiving 50 calls. Quinn *et al.* (2000) compared nursing decisions on the suitability of a patient for transfer out of CCU to a lower dependency ward to patient outcomes, in order to determine if the decision would have been appropriate. This study examined 506 patients, but does not provide details on the number of nurses that took part.

Three of the studies examine nurses' decision-making about treatment for wounds and pressure ulcers (Melchior-MacDougall & Lander 1995, Lamond & Farnell 1998, Letourneau & Jensen 1998). All of the studies use written scenarios that include photographs of the wound as a method of data collection, asking nurses to decide which product they would use. In the studies by Letourneau and Jensen (1998) and Lamond and Farnell (1998), nurses' choices were compared with those of an expert panels. In the study by Melchior-MacDougall and Lander (1995), nurses' choices were evaluated by local nursing practitioners.

Three remaining studies examine nurses' drug administration plans (Corcoran 1986), nurses' ability to identify patients suitable for thrombolytic therapy (Quinn *et al.* 1998), and student nurses' prioritization of nursing interventions (Shamian 1991). All three studies use written case histories or vignettes to collect data on decision-making, and all three use experts as the gold standard against which to compare nurses' decisions.

As in the research into judgement, the study by Quinn *et al.* (2000) has a strength in that it compares nurses' decisions to actual patient outcomes. However, as has already been discussed, this is problematic in decision research because of the nature of uncertainty inherent in such decisions. All of the remaining studies use some form of expert or peer consensus as the gold standard for decision-making. Again, this has strengths in that they provide a standard for decision-making which would generally be seen as matching expert performance (Lipshitz *et al.* 2001). However, all these studies fail to address evaluation of the process of decision-making. Also, by using simulations or vignettes as the basis of data collection, the viewpoint of the person for whom the decisions are being made (the patient) is effectively ignored in the process.

Discussion

The issue of measuring the quality of judgement and decision-making in nursing practice is a complex one. Perhaps because of this complexity, a number of different ways of assessing judgement accuracy and evaluating decisions have been employed (to greater or lesser extents) by researchers examining nurses' judgements and clinical decisions.

Research examining judgement accuracy uses either patient outcomes as the criteria of choice or some form of peer review/expert consensus. Both approaches have their strengths and limitations. Studies using patient outcomes need to accommodate the characteristics of the judgement task within the environment; this includes the predictability of the task in the first place (Engel *et al.* 1990, Cooksey 1996, Dawes 2000, Harries & Harries 2001). Interestingly, existing research tends to focus on examining large numbers of patients, rather than health professionals. In contrast, studies that use expert consensus as the gold standard tend to use simulations or vignettes as the method of data collection, and therefore sample large numbers of professionals rather than patients. Using expert consensus also has its strengths, in that it can provide insight into what is considered to be the standard of clinical practice for that situation. However, as Gonzalez (2001) points out, this may be limiting in that it may be possible to improve on expert performance.

In contrast, the research on nurse decision-making to date has almost exclusively focused on the actual choice made by nurses and compared it with some form of expert or peer review consensus. As has already been highlighted, evaluating decisions by their outcomes is problematic because of the uncertainty inherent in much health care practice. Evaluating decisions by the process by which they are made has been suggested as an alternative approach (Pauker & Pauker 1999), but does not appear to have been employed by nursing researchers to date.

One of the main approaches to measuring the quality of judgements and evaluating decisions in other areas has been through the use of more normative theories, based on probability. This is typified by Bayes' theorem (Hastie & Dawes 2001) for judgements and SEUT (Dowie 1993) for decision-making. None of the research studies we identified used these models as the criterion against which to compare nurses' judgements and decisions. This may be due to the lack of data to provide statistical databases, or nurses' lack of knowledge of such approaches. Although using these types of model as a gold standard for judgements and decisions has been questioned (Jungermann 2000), with the debate mainly surrounding whether or not the 'objective reality' that is suggested by such models actually exists, they can be an important additional tool for measurement in judgement and decision-making research.

Conclusion

The measurement of the quality of judgement and decision-making in nursing practice is highly complicated, and studies that attempt to examine nurses' practice need to acknowledge both the strengths and limitations of the strategies they employ. Existing research into nursing judgements has used simplistic measures of patient outcome as a criterion. Such measures fail to acknowledge the inherent uncertainty in judgement situations, or rely on inter-judge comparisons that can lead to systematic errors.

What is already known about this topic

- Judgement and decision-making in health care are characterized by uncertainty.
- Evaluating judgement and decision-making in practice is complex.
- There are a number of different methods that can be used to assess the quality of judgement and decision-making.

What this paper adds

- A summary of what is known about assessing judgements and decisions in the context of nursing practice.
- A critique of research examining the quality of judgements and decisions.
- Identification of areas where research could be improved in the areas of judgement and decision-making.

Future research needs to use methods that can provide a way of modelling the uncertain nature of judgement tasks, as well as providing an independent criterion measure for accuracy. Approaches such as SJT and Bayes' theorem can provide a framework for this, as well as providing insight into potential areas of error which could then be used as the basis for practice improvement.

Similarly, existing research examining decision-making in nursing uses measures of outcome (without acknowledging the limitations of this approach in terms of the inherent uncertainty in a decision situation), or comparisons with expert panels, which again may be subject to bias. Measuring the goodness of decisions is problematic, and needs an approach that examines both outcomes (which means sampling a large number of decisions) and process. Future research could use the approach of SEUT as a framework for examining the decision process, whilst assesing a large number of decision outcomes. Such techniques offer the promise of being sensitive to complexity and yet able to explicate better the 'goodness' of nurse decision-making and decision-makers in practice.

References

Allen-Davis J.T., Beck A., Parker R., Ellis J.L. & Polley D. (2002) Assessment of vulvovaginal complaints: accuracy of telephone triage and in-office diagnosis. *Obstetrics and Gynecology* **99**, 18–22.

Aspinall M.J. (1979) Use of a decision tree to improve accuracy of diagnosis. *Nursing Research* **28**, 182–185.

Baron J. (2000) Nonconsequentialist decisions. In *Judgment and Decision Making. An Interdisciplinary Reader* (Connolly T., Arkes H.R. & Hammond K.R., eds), Cambridge University Press, Cambridge, UK, pp. 712–732.

Buckingham C.D. & Adams A. (2000) Classifying clinical decision making: a unifying approach. *Journal of Advanced Nursing* **32**, 981–989.

Considine J., Ung L. & Thomas S. (2000) Triage nurses' decisions using the National Triage Scale for Australian emergency departments. *Accident & Emergency Nursing* **8**, 201–209.

Cooksey R.W. (1996) The methodology of social judgement theory. *Thinking and Reasoning* **2**, 141–173.

Corcoran S.A. (1986) Planning by expert and novice nurses in cases of varying complexity. *Research in Nursing & Health* **9**, 155–162.

Dawes R.M. (2000) Proper and improper linear models. In *Judgment and Decision Making. An Interdisciplinary Reader* (Connolly T., Arkes H.R. & Hammond K.R., eds), Cambridge University Press, Cambridge, UK, pp. 378–394.

Department of Health (2000) *The NHS Plan*. HMSO, London.

de Dombal F.T. (1988) Computer-aided diagnosis of acute abdominal pain: the British experience. In *Professional Judgment. A Reader in Clinical Decision Making* (Dowie J. & Elstein A., eds), Cambridge University Press, Cambridge, UK, pp. 190–199.

Dowie J. (1993) Clinical decision analysis: background and introduction. In *Analysing how We Reach Clinical Decisions* (Llewelyn H. & Hopkins A., eds), Royal College of Physicians, London.

Engel J.D., Wigton R., LaDuca A. & Blacklow R.S. (1990) A social judgment theory perspective on clinical problem solving. *Evaluation and the Health Professions* **13**, 63–78.

Goldstein W.M. & Hogarth R.M. (1997) Judgment and decision research: some historical context. In *Research on Judgment and Decision Making. Currents, Connections, and Controversies* (Goldstein W.M. & Hogarth R.M., eds), Cambridge University Press, Cambridge, UK, pp. 3–65.

Gonzalez R. (2001) Decision making in real life. *Journal of Behavioral Decision Making* **14**, 365–366.

Gould D., Kelly D., Goldstone L. & Gammon J. (2001) Examining the validity of pressure ulcer risk assessment scales: developing and using illustrated patient simulations to collect the data. *Journal of Clinical Nursing* **10**, 697–706.

Harries P.A. & Harries C. (2001) Studying clinical reasoning, part 2: applying social judgement theory. *British Journal of Occupational Therapy* **64**, 285–292.

Hastie R. & Dawes R.M. (2001) *Rational choice in an uncertain world: the psychology of judgement and decision making*. Sage, Thousand Oaks CA.

Hastie R. & Rasinski K.A. (1988) The concept of accuracy in social judgment. In *The Social Psychology of Knowledge* (Bar-Tal D. & Kruglanski A.W., eds), Cambridge University Press, Cambridge, UK, pp. 193–208.

Jungermann H. (2000) The two camps on rationality. In *Judgment and Decision Making. An Interdisciplinary Reader* (Connolly T., Arkes H.R. & Hammond, K.R., eds), Cambridge University Press, Cambridge, UK, pp. 575–591.

Kruse J.A., Thill-Baharozian M.C. & Carlson, R.W. (1988) Comparison of clinical assessment with APACHE II for predicting mortality risk in patients admitted to a medical intensive care unit. *JAMA* **260**, 1739–1742.

Lamond D. & Farnell S. (1998) The treatment of pressure sores: a comparison of novice and expert nurses' knowledge, information use and decision accuracy. *Journal of Advanced Nursing* **27**, 280–286.

Lamond D., Crow R., Chase J. & Swinkels K.D.M. (1996) Information sources used in decision making: considerations for simulation development. *International Journal of Nursing Studies* **33**, 47–57.

Leprohon J. & Patel V.L. (1995) Decision-making strategies for telephone triage in emergency medical services. *Medical Decision Making* **15**, 240–253.

Letourneau S. & Jensen L. (1998) Impact of a decision tree on chronic wound care. *Journal of Wound, Ostomy and Continence Nursing* **25**, 240–247.

Lipshitz R., Klein G., Orasanu J. & Salas E. (2001) Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making* **14**, 331–352.

McDonald M., Passik S.D., Dugan W., Rosenfeld B., Theobald D.E. & Edgerton S. (1999) Nurses' recognition of depression in their patients with cancer. *Oncology Nursing Forum* **26**, 593–599.

Marsden J. (2000) An evaluation of the safety and effectiveness of telephone triage as a method of patient prioritization in an ophthalmic accident and emergency service. *Journal of Advanced Nursing* **31**, 401–409.

Maule A.J. (2001) Studying judgement: some comments and suggestions for future research. *Thinking and Reasoning* **7**, 91–102.

Melchior-MacDougall F. & Lander J. (1995) Evaluation of a decision tree for management of chronic wounds. *Journal of Wound, Ostomy and Continence Nursing* **22**, 81–88.

Moore T., Martin J. & Stonehouse J. (1996) Predicting falls: risk assessment tool versus clinical judgement. *Perspectives* **20**, 8–11.

Moyer V.A., Ahn C. & Sneed S. (2000) Accuracy of clinical judgment in neonatal jaundice. *Archives of Pediatric & Adolescent Medicine* **154**, 391–394.

O'Brien J.L., Moser D.K., Riegel B., Frazier S.K., Garvin B.J. & Kim K.A. (2001) Comparison of anxiety assessments between clinicians and patients with acute myocardial infarction in cardiac critical care units. *American Journal of Critical Care* **10**, 97–103.

Papanikolaou P., Clark M. & Lyne P.A. (2002) Improving the accuracy of pressure ulcer risk calculators: some preliminary evidence. *International Journal of Nursing Studies* **39**, 187–194.

Pauker S.P. & Pauker S.G. (1999) What is a good decision? *Effective Clinical Practice* **2**, 194–196.

Quinn T., MacDermott A. & Caunt, J. (1998) Determining patients' suitability for thrombolysis: coronary care nurses' agreement with an expert cardiological 'gold standard' as assessed by clinical and electrocardiographic 'vignettes'. *Intensive and Critical Care Nursing* **14**, 219–224.

Quinn T., Thompson D.R. & Boyle R.M. (2000) Determining chest pain patients' suitability for transfer to a general ward following admission to a cardiac care unit. *Journal of Advanced Nursing* **32**, 310–317.

Reid R.C. & Chappell N.L. (2000) Accuracy of staff assessments in research: dementia and environmental characteristics. *Journal of Mental Health and Aging* **6**, 237–248.

Reischman R.R. & Yarandi H.N. (2002) Critical care cardiovascular nurse expert and novice diagnostic cue utilization. *Journal of Advanced Nursing* **39**, 24–34.

Rosenthal G.E., Mettler G., Pare S., Riegger M., Ward M. & Landefeld S. (1992) Diagnostic judgments of nurse practitioners providing primary gynecologic care: a quantitative analysis. *Journal of General Internal Medicine* **7**, 304–311.

Shamian J. (1991) Effect of teaching decision analysis on student nurses' clinical intervention decision making. *Research in Nursing & Health* **14**, 59–66.

Sox H.C. (1999) What is a good decision? *Effective Clinical Practice* **2**, 196–197.

Tavakoli M., Davies H.T.O. & Thomson R. (1999) Aiding clinical decisions with decision analysis. *Hospital Medicine* **60**, 444–447.

VandenBosch T., Montoye C., Satwicz M., Durkee-Leonard K. & Boylan-Lewis B. (1996) Predictive validity of the Braden scale and nurse perception in identifying pressure ulcer risk. *Applied Nursing Research* **9**, 80–86.

Westfall U.E., Tanner C.A., Putzier D. & Padrick K.P. (1986) Activating clinical inferences: a component of diagnostic reasoning in nursing. *Research in Nursing and Health* **9**, 269–277.

Zou Y., Cole M.G., Primeau F.J., McCusker J., Bellavance F. & Laplante J. (1998) Detection and diagnosis of delirium in the elderly: psychiatrist diagnosis, confusion assessment method, or consensus diagnosis? *International Psychogeriatrics* **10**, 303–308.